P2(b) solution:

1. For feature 1 replace all the ‘?’to ‘b’ which is also the median

For feature 2,3,8,11,14,15 which is real value use the mean of the column to replace the ‘?’

For all the other features use the median of the column

For the real value column we do the Z-scaling to normalize the data

1. See the code

run.py 1 lenses.training lenses.testing

* output : in document testingoflenses1 | perl accuracy.pl

1,1,2,1,3,3

1,1,2,2,1,3

2,1,2,1,3,3

2,2,1,2,2,3

3,1,1,1,3,3

3,2,1,2,2,3

3/6 =0.5

run.py 2 lenses.training lenses.testing | perl accuracy.pl

output: in document testingoflenses2

1,1,2,1,3,3

1,1,2,2,1,3

2,1,2,1,3,3

2,2,1,2,2,3

3,1,1,1,3,3

3,2,1,2,2,3

3/6 = 0.5

/run 1 crx.training.processed crx.testing.processed  | perl accuracy.pl

output in testingofcrx1

/run 3 crx.training.processed crx.testing.processed  | perl accuracy.pl

This part I have a question: I can see that the last 2 columns are different. But the accuracy is always 1. You can check the output in the document, which is right.

For example:



